

ORIGINAL RESEARCH

Price, Sentiment, and Activity in Bitcoin and U.S. Equities: A Comparative Machine-Learning and Predictive-Inference Framework, 2014 to 2026

Arnav Aggarwal

Singapore American School, 40 Woodlands Street 41, Singapore 738547, Singapore

Accepted: October 1, 2026

Abstract

Bitcoin and U.S. equities are forecast in largely separate literatures, which leaves open whether the two markets respond to different forces or whether researchers simply study them differently. This paper imposes one framework on both: four feature blocks (Price/Technical, Sentiment/News, Activity/On-chain-or-Volume, Cross-market) evaluated across six classifier families under a walk-forward design, giving 114 model-block-asset comparisons for Bitcoin (4,346 daily returns) and for Tesla and the S&P 500 (2,338 common-calendar observations each). No comparison exceeds its majority-class base rate by a statistically significant margin after Benjamini-Hochberg correction, so daily directional predictability is not detectable for either asset class. The fitted models nevertheless draw on different information. SHAP attribution assigns 60.3% of Bitcoin's fitted signal to activity and on-chain features and 5.6% to sentiment, against price and technical dominance for Tesla (55.0%) and the S&P 500 (53.5%). Bitcoin's largest single feature is a trading-volume-based illiquidity proxy rather than a blockchain-native metric, so the model's predictions rely on trading activity broadly rather than narrowly on on-chain metrics, and the shares are inconsistent with the claim that cryptoasset prices are driven mainly by hype; because SHAP describes a fitted model rather than the price process, this is evidence about predictive reliance, not causation. A HAC-robust, false-discovery-rate-corrected Granger battery finds that Bitcoin and Tesla returns predict next-day changes in the Crypto Fear & Greed Index ($F = 157.72$ and $F = 15.34$, $p < 0.001$) with no reverse effect. Since volatility and momentum carry half the weight of that index, the Bitcoin result is partly mechanical; the Tesla result cannot be, and is therefore the informative one, although a Granger test establishes predictive precedence rather than the underlying mechanism. GARCH(1,1) estimates place Bitcoin's volatility persistence ($\alpha + \beta = 1.000$) and tail thickness (Student- t $v = 3.20$) above both equities.

Keywords: Bitcoin; Tesla; S&P 500; machine learning; SHAP; sentiment; Granger causality; GARCH

JEL: G12, G14, G17, C45, C58

1. Introduction

Research on equity forecasting has accumulated decades of evidence on what moves prices, from order-flow imbalance and momentum to volatility clustering and news tone, much of it organized around the market-microstructure framework of O'Hara (1995). Cryptoasset research developed along a separate track, and its distinguishing inputs are blockchain-native: active addresses, transaction throughput, and miner revenue have no counterpart in the equity literature. That separation is a research convention, not an empirical finding. Give both asset classes the same predictors and the same evaluation rules, and the question becomes whether they behave like two instances of one market populated with different data, or like two markets with different price-formation mechanics that happen to share a charting interface.

The answer has practical consequences. If Bitcoin's price is driven mainly by sentiment, as the popular claim that crypto runs on hype implies (Beckmann, Geldner, and Wüstenfeld 2024), then social-media mood and news tone should be the primary crypto risk-surveillance inputs, and equity-style technical analysis should transfer poorly to digital assets. If instead Bitcoin's price formation rests on the same real-activity signals that underpin valuation elsewhere, then the appropriate toolkit lies closer to on-chain fundamentals than to sentiment indices, and cross-asset portfolio construction should treat Bitcoin as a structurally distinct exposure rather than as a volatile equity.

This paper builds one four-block feature architecture, Price/Technical (P), Sentiment/News (S), Activity/On-chain-or-Volume (A), and Cross-market (X), for Bitcoin, Tesla (TSLA), and the S&P 500 (SPY) simultaneously. All eight block combinations (P, S, A, PS, PA, SA, PSA, PSAX) are evaluated across six model families in a walk-forward out-of-sample design, SHAP attribution decomposes what each asset's fitted model rests on, and a HAC-robust, false-discovery-rate-corrected Granger battery tests whether returns lead crypto-market sentiment or follow it.

Three contributions follow. The first is methodological: existing studies commit to one asset class and one modelling family, which makes claims about what drives "the market" incomparable across assets. The grid here yields 114 comparisons under one protocol, with the multiple-comparison correction applied to the full family rather than to the surviving cell. The second is interpretive: rather than assuming that cryptoassets are sentiment-driven, the study measures the composition of predictive signal from fitted models, turning the hype hypothesis into a testable claim about block shares. The third concerns predictive direction: composite sentiment indices are built in part from realized price behavior, so a Granger design read against the index's published construction weights separates a mechanical channel from a potentially behavioral one, while establishing predictive precedence rather than an underlying mechanism. Section 2 reviews the literature and states the hypotheses, Section 3 sets out data and method, Section 4 reports results, Sections 5 to 7 cover robustness, conclusions, and limitations.

2. Related literature and hypotheses

2.1 Return predictability, sentiment, and on-chain fundamentals

Fama's (1970) efficient-market benchmark holds that prices incorporate available information, implying that no predictor should earn systematic out-of-sample profit. The weak form restricts this to past prices alone, and the momentum literature reported apparent violations: Jegadeesh and Titman (1993) found that stocks ranked on three- to twelve-month past returns continued to outperform over subsequent holding periods, a result robust across markets and construction choices. That tension between the weak form and observed momentum is why a price and technical block belongs in any predictability test.

Baker and Wurgler (2006) showed that investor sentiment predicts cross-sectional equity returns mainly among stocks that are hard to value and costly to arbitrage, a description that travels well to an asset with no cash flows or pricing history. Tetlock (2007) moved the measurement problem into text and found that media pessimism predicts short-run price declines and heavier volume, establishing that tone carries information distinct from the events it describes. A parallel strand measures attention rather than tone: Kristoufek (2013) related Bitcoin prices to search and Wikipedia activity and found dependence running both ways, an early warning that attention measures built from observable behavior can be endogenous to price. Composite gauges followed. The Crypto Fear & Greed Index combines six weighted inputs, with volatility and market momentum/volume at 25% each, social-media activity and surveys at 15% each, and Bitcoin dominance and Google Trends at 10% each (Alternative.me 2024). Two features of that construction matter downstream: the index is market-wide but heavily Bitcoin-weighted, and at least half of it is computed from realized price behavior, so any association with recent or contemporaneous returns contains a mechanical component that a causal reading must isolate. The equity counterpart used here, the CBOE volatility index, is documented by Whaley (2000).

A separate literature treats network usage as the fundamental of a blockchain asset. Alabi (2017) fitted Metcalfe-type power laws to the user bases of Bitcoin, Ethereum, and Dash and found network value tracking active users closely, which supports the argument that value derives from the network that settles transactions. Practitioners operationalized this into the network-value-to-transactions (NVT) ratio, miner-revenue indicators such as the Puell Multiple, and hash-ribbon capitulation signals, all of which enter the Activity block for Bitcoin here. No equity issuer maintains a comparable ledger, and that is the weakest joint in any cross-asset design claiming a fixed predictor menu. The Activity block for Tesla and the S&P 500 therefore uses the closest daily-frequency real-activity signals available: dollar volume, the Amihud illiquidity measure, on-balance-volume slope, and the volume-price trend.

2.2 Machine learning in financial forecasting

Gradient-boosted trees became the default for structured prediction once implementations reached the scale of tabular financial data. Chen and Guestrin (2016) introduced XGBoost, and Ke et al. (2017) developed LightGBM, which reduces training cost sharply through histogram binning and leaf-wise growth. Recurrent architectures address sequential dependence: Hochreiter and Schmidhuber (1997) introduced the long short-term memory cell to solve the vanishing-gradient problem, and Cho et al. (2014) proposed the gated recurrent unit as a lighter variant. Fischer and Krauss (2018) applied LSTM networks to S&P 500 constituent direction and reported returns exceeding standard benchmarks, though their design differs from a single-asset daily ablation in ways that limit comparison: they traded a broad cross-section, used relative return ranks as the target instead of a majority-class benchmark, and reported pre-cost results. Surveys of this literature cover one asset class each, Khedr et al. (2021) for crypto and Nabipour et al. (2020) for equities, which is precisely the separation this paper tests instead of assuming.

2.3 Cross-market spillovers and the research gap

Bouri et al. (2017) used dynamic conditional correlations to show that Bitcoin diversifies rather than hedges against equities, bonds, oil, gold, and the dollar, with safe-haven properties confined to specific horizons and markets, while a later survey of the trading literature documents the market-structure, liquidity, and integrity risks that surround the asset class (Fang et al. 2022). Within crypto, Shen, Urquhart, and Wang (2019) found that tweet counts predict next-day volume and realized volatility under both linear

and nonlinear tests, and Da Gama Silva et al. (2019) documented herding and contagion among the most liquid coins. Three gaps remain. Cross-asset claims usually rest on studies with different samples, targets, and evaluation rules, so comparability is untested rather than established. Inference is exposed to data snooping, since papers report the best cell from a large implicit search space without correcting for it (White 2000; Benjamini and Hochberg 1995). And no study decomposes feature-block attribution across asset classes to test whether models of crypto prices really do rely disproportionately on sentiment. Three hypotheses follow.

- **H1 (predictability).** After false-discovery-rate correction across the full grid, no cell exceeds its majority-class base rate by a statistically significant margin at daily horizon.
- **H2 (attribution composition).** Bitcoin's fitted model places its largest block share on Activity, Tesla and the S&P 500 place theirs on Price/Technical, and sentiment accounts for a smaller share for Bitcoin than for either equity.
- **H3 (predictive direction).** Returns lead changes in crypto-market sentiment. The asymmetry is expected to be larger for Bitcoin, because the sentiment index is built partly from Bitcoin's own volatility and momentum, a channel that does not exist for Tesla.

3. Data and Methods

3.1 Sample

Bitcoin trades continuously and yields 4,347 daily price observations, hence 4,346 log returns, from 17 September 2014 to 11 August 2026. Tesla and the S&P 500 exchange-traded fund trade on the American equity calendar, giving roughly 2,990 trading days over the same window. Three sample concepts are kept separate throughout: the *available* sample is the set of days on which an asset traded; the *common complete-case* sample is the subset on which every column entering a given table is simultaneously observed, which is 2,338 days for Tesla and the S&P 500 and is the basis of Tables 1 and 3; and per-cell samples in the ablation grid depend on the missingness pattern of that block and are reported in Table 4. Conflating these counts is a common source of arithmetic inconsistency in multi-asset panels.

Price data are merged with four families of conditioning variables: sentiment and news (the Crypto Fear & Greed Index and the CBOE volatility index, VIX); activity (Bitcoin on-chain metrics, namely active addresses, transaction count and dollar volume, hash rate, miner revenue, fees, average block size, mempool size, difficulty, the NVT ratio, the Puell Multiple, and a hash-ribbon signal, plus market-microstructure volume measures for all three assets); and cross-market series (lagged returns, volatility, absolute returns, and momentum of the other two assets). A GDELT-derived news-tone series is retained in the merged panel as a supplementary candidate feature but is not part of the Cross-market block as specified below. All series are aligned to a common daily calendar and merged into one 315-column panel, and each asset-block combination is estimated on its largest complete-case subsample. Complete-case handling was adopted in place of imputation for comparability with earlier multi-methodology forecasting work; its cost is quantified in Section 7.

3.2 Feature blocks

Four blocks are built with identical internal structure across assets, differing in column content where the asset class requires it. The blocks are disjoint, so no predictor appears in more than one, which is what

allows incremental contribution to be read from the grid.

- **Price/Technical (P):** returns at horizons of 1 to 63 days; realized and downside volatility at several windows; skewness and kurtosis; price-to-moving-average ratios (5, 10, 20, 50, 200 days); RSI (7, 14); MACD and histogram; stochastic %K; Bollinger Band position and width; normalized ATR; candle features (high-low range, close-to-open gap, intraday range, close location); 52-week position; drawdown from peak; streak indicators.
- **Sentiment/News (S):** the Crypto Fear & Greed Index level, 1-day and 7-day changes, 30-day z-score, extreme-fear and extreme-greed flags, and 7-day moving average; VIX level, 1-day change, 30-day z-score, and term-structure proxy.
- **Activity/On-chain-or-Volume (A):** for Bitcoin, the on-chain usage and miner-economics metrics of Section 3.1 alongside market-microstructure volume proxies (dollar volume, Amihud illiquidity, on-balance-volume slope, volume-price trend); for Tesla and the S&P 500, the volume proxies alone, standing in for real activity for the reasons given in Section 2.1.
- **Cross-market (X):** 1-day and 5-day lagged returns, 21-day realized volatility, absolute return, and 21-day momentum of the other two assets. Crypto sentiment enters through the Sentiment block only, so X measures price spillover without duplicating a sentiment input.

The eight combinations are P, S, A, PS, PA, SA, PSA, and PSAX. The target in every case is the binary direction of the asset's own next-day return.

3.3 Models, attribution, and causal tests

Augmented Dickey-Fuller (ADF) and KPSS tests are applied to each return series, with lag order by the Akaike criterion and automatic bandwidth respectively. A GARCH(1,1) model with Student-*t* innovations is estimated by quasi-maximum likelihood (Engle 1982; Bollerslev 1986) with returns in percent, so that the estimated variance is in percent squared; these estimates are descriptive and are not inputs to the classifiers, which take realized volatility through block P. Four classical classifiers (logistic regression, random forest, XGBoost, LightGBM) are estimated on all eight blocks, and two recurrent models (single-layer LSTM and GRU, 32 hidden units, 20% dropout, 10-day lookback, Adam) on the P, PSA, and PSAX blocks only, a scope restriction dictated by grid size: 32 cells per asset plus six gives 38, and 114 across the three assets. SHAP values (Lundberg and Lee 2017) are computed with the tree explainer on an XGBoost model fitted to PSAX for each asset and aggregated into block shares of total mean absolute SHAP value. Granger causality (Granger 1969) is tested as Wald tests on one-lag autoregressive distributed-lag regressions with Newey and West (1987) HAC covariance at five lags, across 18 directional pairs spanning returns, realized volatility, sentiment changes, and on-chain activity growth. Finally, walk-forward equity curves are constructed for PSAX under XGBoost and logistic regression, long or short one unit according to predicted direction, net of a 10-basis-point round-trip cost, against buy-and-hold.

3.4 Evaluation protocol

For each asset-block combination the complete-case sample is split chronologically 80/20. The final 20% is divided into five approximately equal expanding-window folds; at the start of each fold the model is refit on all preceding data, predictions are generated for the fold, and predictions are pooled across folds

to compute directional accuracy. Every model is thus refit five times, so accuracy cannot be an artifact of one favorable split. Two features of the protocol require comment. The base rate is the majority-class share within the pooled test region, which compares the model against an oracle that knows test-period class balance; it is a demanding benchmark, but since the base rate is a realized frequency and not an estimated proportion, the one-sided two-proportion z -test used here is an approximation, and Section 5 reports an exact binomial alternative. Significance is assessed per cell and then corrected with the Benjamini and Hochberg (1995) procedure within each asset across all 38 cells, giving three families of 38 and 114 tests in total; correcting across the pooled 114 would be stricter and cannot change the conclusions. Hyperparameters are fixed at moderate values rather than tuned per cell, a simplification whose implications Section 7 treats as a limitation rather than a claim about model capacity.

3.5 Analytical framework

Consider an asset i traded by three participant types, momentum (M), sentiment (S), and activity fundamentalists (A), with population mass shares $w_i^T \geq 0$ summing to one. Each type trades on its own lagged signal, $M_{i,t-1}$, $S_{i,t-1}$, and $A_{i,t-1}$, where the activity signal is on-chain usage for Bitcoin and a volume-based liquidity proxy for the equities. Market clearing under linear price impact gives

$$\Delta P_{i,t} = w_i^M \beta^M M_{i,t-1} + w_i^S \beta^S S_{i,t-1} + w_i^A \beta^A A_{i,t-1} + \varepsilon_{i,t} \quad (1)$$

with $\beta^M, \beta^S, \beta^A > 0$ taken as common across assets and $\varepsilon_{i,t}$ idiosyncratic noise. Because Bitcoin's holder base carries a comparatively large mass of participants whose activity is observable on the blockchain (miners, large holders, long-term users), while Tesla and the S&P 500 are dominated by participants trading on price statistics and flow, $\partial^2 \Delta P_{i,t} / \partial A_{i,t-1} \partial w_i^A = \beta^A > 0$: an asset with more activity traders should load more heavily on block A, and, holding w_i^M fixed, less heavily on sentiment. Writing σ_k for the standard deviation of block- k signals, the predicted block share is

$$S_i^k = \frac{w_i^k \beta^k \sigma_k}{\sum_{j \in \{M, S, A\}} w_i^j \beta^j \sigma_j} \quad (2)$$

which is what the SHAP decomposition in Section 4.4 estimates. Two caveats discipline that reading. The mapping from (2) to a SHAP share requires that fitted models rank features as the population does, that correlated predictors divide credit as the Shapley convention dictates, which matters because on-chain and trading volume are related, and that a model trained for classification error nonetheless weights signals carrying genuine demand. Equation 2 therefore motivates the comparison across assets without identifying the underlying parameters: the claim tested below is comparative, whether block shares are ordered differently for Bitcoin than for the equities, not whether a share equals $w_i^k \beta^k \sigma_k$. The model is also silent on forecastability, since (1) describes contemporaneous impact; whether lagged signals predict return direction out of sample is settled empirically in Section 4.3.

The framework has a second, separate implication for predictive direction. The sentiment index is assembled from market observables, so writing it as $F_t = g(\cdot)$ with $\partial g / \partial r_t \neq 0$, where r_t is the current Bitcoin return, gives $\text{Cov}(r_{t-1}, \Delta F_t) \neq 0$ from the construction alone. A finding that returns predict next-day index changes therefore mixes a mechanical channel with any behavioral feedback, and the test alone cannot separate them. The construction weights do allow separation across assets. Bitcoin enters the index directly through its volatility, momentum, volume, dominance, and search interest, so a

positive finding for Bitcoin is partly definitional. Tesla enters no component, so a Tesla finding cannot be mechanical and could reflect a common factor linking equity risk appetite to crypto conditions, or a spillover from equity price action to crypto sentiment, although a Granger test establishes only predictive precedence and cannot tell these mechanisms apart. That differential prediction is H3.

4. Empirical results

4.1 Descriptive statistics and stationarity

Bitcoin earned the highest mean daily return in the sample, 0.114%, and carried the highest volatility, with negative skewness and excess kurtosis of 11.9 (Table 1). The combination of positive mean and fat left tail is the profile of an asset that compounds upward while suffering repeated severe drawdowns, and Figure 1 shows how deep those were: Bitcoin's peak-to-trough decline approached 80% during 2018 and 2019, far beyond anything the S&P 500 experienced. Tesla's mean return is close to zero despite volatility of 53.9%, comparable to Bitcoin's 55.5%, so the gap between the two lies in central tendency, not dispersion. The S&P 500 shows materially lower volatility at 17.5% but still carries excess kurtosis of 11.5, almost entirely attributable to a few crash episodes, the COVID-19 shock of early 2020 above all. Annualized figures use the trading-day convention ($\bar{r} \times 252$ and $s \times \sqrt{252}$) for all three assets so that they are comparable; a 365-day convention would raise Bitcoin's annualized mean to 41.6% and volatility to 66.7%.

Table 1. Descriptive statistics for daily log returns. Bitcoin covers 17 September 2014 to 11 August 2026; Tesla and the S&P 500 cover the common complete-case sample of 2,338 trading days over the same window.

Asset	<i>N</i>	Mean (%/day)	SD (%/day)	Skew- ness	Excess kurt.	Ann. return (%)	Ann. vol. (%)
BTC	4,346	0.114	3.493	−0.708	11.94	28.68	55.45
TSLA	2,338	0.002	3.396	−0.019	4.32	0.51	53.92
SPY	2,338	0.044	1.099	−0.090	11.51	11.11	17.45

Figure 2 adds three views. The return distributions make the tail difference visible: Bitcoin and Tesla share a similar dispersion and both are far wider than the index. Twenty-one-day realized volatility for Bitcoin and Tesla occupies a common 25% to 100% range for most of the sample, with Bitcoin's excursions well above it in 2018 and 2020. The 126-day rolling correlations place Tesla and the S&P 500 highest throughout, Bitcoin and Tesla lowest, occasionally negative in 2016 and 2017, with all three pairs trending upward after 2020. That pattern is consistent with gradual rather than permanent integration of crypto into the broader risk complex. Table 2 confirms that all three return series are stationary: ADF rejects a unit root at 1% and KPSS fails to reject stationarity at 10%, so returns are treated as integrated of order zero and used without differencing.

Table 2. Panel A: stationarity tests on daily returns, ADF null is a unit root and KPSS null is stationarity. Panel B: GARCH(1,1) estimates with Student- t innovations by quasi-maximum likelihood, returns in percent, so ω is in percent squared. Information criteria are not comparable across assets because the series differ in length.

Panel A				
Asset	ADF stat.	ADF p	KPSS stat.	KPSS p
BTC	−20.050	< 0.001	0.166	> 0.100
TSLA	−48.021	< 0.001	0.044	> 0.100
SPY	−14.638	< 0.001	0.022	> 0.100

Panel B					
Asset	α	β	$\alpha + \beta$	ν	AIC
BTC	0.108	0.892	1.000	3.20	21304.7
TSLA	0.034	0.962	0.996	4.01	11840.7
SPY	0.200	0.789	0.990	5.45	5971.8

[Figure 1. Cumulative growth of one dollar on a log scale (panel A) and drawdown from running peak (panel B), 2014 to 2026. Bitcoin delivers the largest cumulative return and the deepest drawdowns, approaching 80% in 2018 and 2019.]

[Figure 2. (A) Daily return distributions. (B) 21-day realized volatility, annualized. (C) 126-day rolling pairwise correlations; Tesla to S&P 500 is consistently highest, Bitcoin to Tesla consistently lowest.]

4.2 Volatility dynamics and cross-asset correlation

Panel B of Table 2 reports the GARCH estimates. Bitcoin's persistence, $\alpha + \beta = 1.000$, sits at the boundary of covariance stationarity and implies volatility shocks that never fully decay, while Tesla at 0.996 and the S&P 500 at 0.990 remain close to it. The estimated degrees of freedom reproduce the tail ordering of Table 1, with Bitcoin's $\nu = 3.20$ the lowest and therefore the heaviest tailed. Two caveats apply: information criteria are reported for completeness but are not comparable across rows, and the estimates are descriptive, since volatility enters the classifiers through realized measures in block P rather than through the fitted conditional variance. Figure 3 shows the fitted series, where clustering around the 2018 and 2020 stress episodes is visible in all three assets and most extreme for Bitcoin, whose annualized conditional volatility exceeded 200% in March 2020.

[Figure 3. GARCH(1,1) annualized conditional volatility, 2014 to 2026. Bitcoin's is consistently the highest, with the most pronounced volatility of volatility; all three cluster around the 2018, 2020, and 2022 stress episodes.]

Table 3 reports the correlation structure. The Tesla to S&P 500 correlation of 0.491 is roughly double the Bitcoin to S&P 500 value of 0.251 and well above Bitcoin to Tesla at 0.187, so despite a superficially equity-like price path, Bitcoin remains considerably less integrated with the broad market than one large-cap stock is with another. The VIX correlations carry the more informative contrast: −0.754 with the S&P 500 and −0.407 with Tesla are both far larger in magnitude than the −0.203 with Bitcoin, indicating comparative insulation from the equity market's principal fear gauge. The final row is equally telling in the other direction, since daily changes in crypto sentiment are essentially uncorrelated with same-day returns for all three assets, at absolute values of 0.023 or below. A sentiment measure uncorrelated with

contemporaneous returns yet predicted by past returns, as Section 4.5 shows, is a summary of recent price behavior rather than a leading indicator.

Table 3. Pairwise Pearson correlations among daily returns and levels on the common complete-case sample ($N = 2,338$). $r_{\cdot}$ denotes daily return, d_{fng} the daily change in the Crypto Fear & Greed Index. Lower triangle shown.

	r_{BTC}	r_{TSLA}	r_{SPY}	r_{VIX}	r_{GOLD}	r_{DXY}	d_{fng}
r_{BTC}	1.000						
r_{TSLA}	0.187	1.000					
r_{SPY}	0.251	0.491	1.000				
r_{VIX}	-0.203	-0.407	-0.754	1.000			
r_{GOLD}	0.088	0.064	0.065	-0.002	1.000		
r_{DXY}	-0.087	-0.078	-0.116	0.045	-0.392	1.000	
d_{fng}	0.023	0.015	0.004	0.006	0.017	-0.004	1.000

4.3 Directional accuracy

Table 4 and Figure 4 report the central predictability test. Across the 114 cells the largest raw accuracy gain over the base rate is 2.93 percentage points (Tesla, Activity-only, LightGBM) and the largest shortfall is 9.22 points (S&P 500, Sentiment-only, LightGBM). Figure 4 averages across models within each block and shows the dominant pattern: negative for almost every block and asset, with the single positive bar belonging to Tesla's Activity-only block. The table reports the largest accuracy gain per asset, which is not the cell with the smallest p -value, since a smaller gain measured on a larger complete-case sample can be more precisely estimated than a larger gain on a small one. The final column reports the smallest FDR-adjusted p -value across that asset's 38-cell family, which belongs to a different cell from the one shown. Under correction at 5%, those family minima are 0.938 for Bitcoin, 0.916 for Tesla, and 0.998 for the S&P 500, so no cell for any asset, block, or model is distinguishable from its base rate. H_1 is supported. Several cells do exceed their base rate in raw accuracy, and the Tesla cell exceeds it by nearly three percentage points; what fails is the inference, not the sign. Once 38 tests per asset are accounted for, none of those gains survives sampling variation. This contrasts with Fischer and Krauss (2018), who reported profitable LSTM-based S&P 500 direction prediction; their design traded a broad cross-section under a different target definition and without a multiple-comparison correction across model and feature choices, so the two results are not directly comparable, but the difference is large enough to note.

Table 4. Largest accuracy gain over the majority-class base rate per asset, out of 38 model-block combinations per asset and 114 in total. The final column reports p_{min} , the smallest Benjamini-Hochberg adjusted p -value in that asset's family, which belongs to a different cell. No cell survives correction at 5%.

Asset	Best block	Best model	n (test)	Base rate	Accuracy	p_{min} (FDR)
BTC	PSAX	LSTM	614	0.508	0.534 (+2.61 pp)	0.938
TSLA	A	LightGBM	581	0.509	0.539 (+2.93 pp)	0.916
SPY	P	Random forest	548	0.569	0.569 (+0.00 pp)	0.998

[Figure 4. Mean out-of-sample accuracy minus majority-class base rate, in percentage points, averaged

across the models estimated on each block. No combination survives Benjamini-Hochberg correction at 5% for any asset.]

4.4 SHAP block attribution

The absence of exploitable predictability does not imply that the fitted models rely on the same information; it implies that whatever structure exists is not exploitable at daily out-of-sample horizons with these methods. Attribution reveals a sharp divergence (Figure 5). Bitcoin's fitted model places 60.3% of total mean absolute SHAP value on Activity/On-chain-or-Volume, then 24.8% on Price/Technical, 9.3% on Cross-market, and 5.6% on Sentiment/News. Tesla and the S&P 500 invert the ordering, with Price/Technical dominant at 55.0% and 53.5%, Activity a distant second at 21.5% and 13.0%, and sentiment higher than Bitcoin's at 9.1% and 12.2%. H2 is supported in its main ranking. On this evidence the fitted models offer no support for the claim that cryptoasset prices are driven disproportionately by hype, since Bitcoin's model leans on sentiment least of the three assets. SHAP values describe how a particular fitted model forms its predictions, not how prices are generated, so these shares are evidence about predictive reliance rather than a causal finding about price formation.

Table 5 lists the twelve highest-attribution features for Bitcoin, nine from the Activity block, led by the Amihud illiquidity proxy computed from Bitcoin's own trading volume at 11.61% on its own, with the network-value-to-transactions measure fifth at 2.40%. The composition of the block matters. Of the nine Activity features in the top twelve, the three market-microstructure volume proxies account for 16.9 percentage points between them, against 10.3 points for the five blockchain-native features. The largest individual feature in the model is therefore a trading-volume measure, not a native on-chain quantity, which makes Bitcoin's model predictions better described as relying on activity in a broad sense, trading intensity as well as network usage, than on narrowly on-chain metrics. Whether Bitcoin's price formation itself rests on these signals remains a hypothesis that attribution alone cannot establish. One limitation should be stated plainly: shares are point estimates from a single fitted model per asset, and the gaps carrying the interpretation, 60.3% against 55.0%, and 5.6% against 9.1% and 12.2%, have no confidence intervals attached.

[Figure 5. Share of total mean absolute SHAP value by feature block, PSAX XGBoost model, held-out test period. Bitcoin is Activity-dominated at 60.3%; Tesla and the S&P 500 are Price/Technical-dominated at 55.0% and 53.5%.]

Table 5. Twelve highest-attribution features for Bitcoin, PSAX XGBoost model, as percentages of total mean absolute SHAP value. Nine of the twelve belong to the Activity block.

Rank	Feature	Share	Rank	Feature	Share
1	A_mkt_amihud	11.61	7	A_mkt_dlogvol	2.15
2	P_ret_lag3	3.66	8	A_avg_block_size_d7	2.14
3	A_mkt_dollar_vol	3.16	9	A_tx_volume_usd_d7	1.84
4	X_spy_abs_ret	2.86	10	P_ret_5	1.84
5	A_nvt	2.40	11	P_vol_10	1.77
6	A_tx_volume_usd_re130	2.17	12	A_active_addresses_d7	1.76

4.5 Trading validation and causality

Figure 6 and Table 6 report walk-forward equity curves for the PSAX specifications net of a 10-basis-point round-trip cost. The results are asset- and window-dependent, not uniformly positive. The Bitcoin strategy gains in a test window dominated by a sustained drawdown, returning 34% to 39% against a buy-and-hold loss of 40%, while the S&P 500 strategy loses 9% to 37% against a buy-and-hold gain of 23%. Tesla's strategy tracks its benchmark at -9% to -12% against -12% . That pattern of large, sign-flipping differences driven by which regime the test window contains is what the null in Section 4.3 predicts: a genuine and stable edge would show consistent gains net of costs across assets, and it does not. Two conventions apply: Sharpe ratios are annualized from daily arithmetic returns by $\sqrt{252}$, and their positive signs for Tesla coexist with a negative compound return because the arithmetic mean can remain positive while volatility drag pulls terminal wealth below one. Costs are charged at a conservative 10 basis points round trip, which is generous to the strategies in the equity cases, since the model trades daily.

[Figure 6. Out-of-sample equity curves, PSAX XGBoost strategy against buy-and-hold over the walk-forward test period, net of a 10-basis-point round-trip cost. Bitcoin gains during a drawdown window, Tesla tracks its benchmark, and the S&P 500 loses during a rally.]

Table 6. Walk-forward strategy performance on PSAX, net of a 10-basis-point round-trip cost, against buy-and-hold. SR is the annualized Sharpe ratio; DD is maximum drawdown in percent.

Asset	Model	Strat. / B&H return (%)	Strat. / B&H SR	Strat. / B&H max DD (%)
BTC	Logistic	+38.8 / -40.0	+0.55 / -0.39	-31.9 / -56.8
BTC	XGBoost	+34.3 / -40.0	+0.51 / -0.39	-39.2 / -56.8
TSLA	Logistic	-8.6 / -12.3	+0.19 / +0.15	-56.8 / -57.6
TSLA	XGBoost	-12.0 / -12.3	+0.15 / +0.15	-52.8 / -57.6
SPY	Logistic	-8.7 / +22.9	-0.23 / +0.81	-20.8 / -19.5
SPY	XGBoost	-36.5 / +22.9	-1.47 / +0.81	-38.8 / -19.5

Table 7 reports the HAC-robust Granger battery. After correction, exactly two of eighteen relationships survive at 5%, both in the same direction: Bitcoin returns Granger-cause next-day changes in the Crypto Fear & Greed Index ($F = 157.72$, adjusted $p < 0.001$) and Tesla returns do the same ($F = 15.34$, adjusted $p < 0.001$). The reverse direction is insignificant, with raw $p = 0.482$ and $p = 0.572$ (0.748 for both after correction, as in Table 7). No on-chain activity growth measure forecasts Bitcoin returns, and no Bitcoin to S&P 500 volatility spillover survives. H3 is supported in direction and asymmetry, but the two surviving results are not equally informative. The Bitcoin finding is partly mechanical, since volatility and momentum each carry 25% of the index's weight and dominance a further 10%, so at least half of the index is computed from recent price behavior; a test that returns predict next-day index changes is in part a test of the index's own definition, and the very large F of 157.72 fits that reading rather than contradicting it. What it establishes is that the index lags Bitcoin price behavior, which matters because it is widely used as if it led. The Tesla finding cannot be explained that way, since Tesla enters no component of the index; an F of 15.34 with no reverse effect could reflect a common factor linking equity risk appetite to crypto conditions, or a spillover from equity price action to crypto sentiment. The test establishes predictive precedence, that past Tesla returns carry information about next-day index changes, but it cannot distinguish between these mechanisms or identify the underlying economic channel. That is the

more informative result, and the fact that the asymmetry appears for Tesla and not for the S&P 500 is consistent with Tesla acting as a high- β risk-appetite proxy among the assets available here. The practical implication is narrow but firm: using the index as an early-warning indicator for future Bitcoin moves is not supported, and past index changes carry no significant information about subsequent returns in either asset. The result sits alongside Shen, Urquhart, and Wang (2019), who found that attention predicts next-day volume and volatility: attention and trading activity appear to move together, while predictive precedence in the sentiment-price relationship runs from price to the index rather than back.

[Figure 7. Granger causality across 18 directional pairs, ranked by $-\log_{10}(p)$ from HAC-robust Wald tests. Dark bars survive Benjamini-Hochberg correction at 5%; the Bitcoin to sentiment and Tesla to sentiment pairs are the only two that do.]

Table 7. HAC-robust Granger-causality Wald tests, Newey-West standard errors at five lags, 18 directional pairs, Benjamini-Hochberg corrected within the set of 18. Only the two pairs in the top block survive at 5%.

Cause	Effect	F	p (FDR)	Cause	Effect	F	p (FDR)
r_{BTC}	d_{fng}	157.72	< 0.001	d_{fng}	r_{BTC}	0.49	0.748
r_{TSLA}	d_{fng}	15.34	< 0.001	rv_{SPY}	rv_{BTC}	0.41	0.748
rv_{BTC}	rv_{SPY}	6.27	0.074	d_{fng}	r_{TSLA}	0.32	0.748
r_{TSLA}	r_{BTC}	4.36	0.166	$g_{BTC,tx}$	r_{BTC}	0.29	0.748
r_{TSLA}	r_{SPY}	3.12	0.270	rv_{VIX}	r_{BTC}	0.19	0.748
rv_{VIX}	r_{TSLA}	2.82	0.270	rv_{TSLA}	rv_{BTC}	0.19	0.748
$g_{BTC,active}$	r_{TSLA}	2.63	0.270	r_{SPY}	r_{TSLA}	0.19	0.748
r_{BTC}	r_{TSLA}	1.89	0.382	$g_{BTC,txvol}$	r_{BTC}	0.07	0.835
rv_{BTC}	rv_{TSLA}	0.52	0.748	$g_{BTC,active}$	r_{BTC}	0.00	0.986

5. Robustness

Four checks bear on the interpretation. First, the protocol refits every model five times on an expanding window, so accuracy cannot be an artifact of one favorable split. Second, the correction is applied to the disclosed family rather than to the surviving cell, and the family minimum is reported beside the best cell so that the distance from significance is visible; pooling all 114 cells into one family would be stricter and cannot change the result, since multiplying a raw p -value by a larger family size only increases adjusted values that already exceed 0.9. Third, the base-rate benchmark was re-examined with an exact one-sided binomial test against the same fixed rate, which gives $p \approx 0.08$ for the largest-gain cell in place of the 0.159 from the two-proportion z -test; multiplied by 38 this still exceeds one, and every other cell has a smaller gain or a smaller sample, so the null is unchanged. Fourth, the trading exercise is an economic check independent of the statistical test, and its value lies in the comparison it fails to support: a concealed exploitable signal would show consistent net-of-cost gains across assets, while the observed pattern is regime-dependent and sign-flipping.

Two standard forecast-comparison procedures are not applied here and are the natural next step. Diebold and Mariano (1995) test whether the accuracy of one forecast sequence differs from another, and Hansen (2005) extends that logic to many competing models while controlling the familywise error rate. Neither is the primary tool in this design, since the benchmark is a fixed classification rule and not a rival forecast, and both would matter in a follow-up that pits the block-based models against a specific competing specification.

Two internal-consistency observations reinforce the picture. The attribution of Section 4.4 and the correlations of Section 4.2 were computed independently, from a fitted model and from raw moments, and agree that Bitcoin's returns and fitted model are less tied to the conventional fear-gauge channel. The causality results agree with the near-zero contemporaneous correlations of Table 3: an index whose daily changes are uncorrelated with same-day returns yet predictable from past returns behaves as a lagging summary of price action. One near-threshold result deserves mention so that it is not over-read. The Bitcoin to S&P 500 volatility spillover has a raw p of 0.012 and an adjusted value of 0.074. It fails correction, which is precisely what correction exists to enforce for a third-ranked test in a family of eighteen, and it belongs in a dedicated volatility study where it is the object of interest rather than one row in a battery.

6. Discussion and conclusion

The paper asked whether Bitcoin and U.S. equities, given identical predictors and evaluation rules, behave like one kind of market running on different data or like two markets with different price-formation mechanics. Their fitted models turn out to differ in the composition of their predictive signal but not, at this horizon, in exploitability.

On exploitability the result is negative and unambiguous. Across 114 model-block-asset combinations, no specification exceeds its majority-class base rate by a statistically significant margin once the family of tests is corrected, under a protocol built to survive the data-snooping critique that weakens single-cell claims. Daily-frequency price, sentiment, and activity data, however comprehensively engineered, did not produce a one-day-ahead edge for any of the three assets.

On composition the result runs against the prevailing narrative. Attribution places 60.3% of Bitcoin's fitted signal on activity and on-chain features and 5.6% on sentiment, against price and technical dominance for Tesla and the S&P 500 at 55.0% and 53.5%. Bitcoin's largest single feature is a trading-volume-based illiquidity proxy, not a blockchain-native metric, which tempers the on-chain reading without overturning it: what Bitcoin's fitted model relies on is real usage and trading intensity, understood broadly, and not mood. That price formation itself works this way is a hypothesis the attribution is consistent with, not a causal finding it establishes. The claim that cryptoasset prices are driven disproportionately by sentiment finds no support in the fitted models, and on the block-share measure the ordering runs the other way.

On predictive direction the two surviving results separate a definitional channel from a behavioral one. Bitcoin and Tesla returns both predict next-day changes in the Crypto Fear & Greed Index with no reverse effect, but at least half of the index is constructed from realized price behavior, so the Bitcoin finding is substantially mechanical. The Tesla finding cannot be, since Tesla enters no component of the index, and it is therefore the evidence that matters: Tesla returns have predictive precedence over crypto-market sentiment, which could reflect equity risk appetite leading crypto sentiment, although the Granger test cannot identify that mechanism. Three implications follow. Traders should treat daily sentiment indices as descriptive of recent conditions rather than predictive of near-term direction. Crypto surveillance should weight activity and volume measures, liquidity proxies such as the Amihud measure among them, more heavily than social-media tone, while noting that the dominant individual signal in this study is a market-volume proxy rather than a purely on-chain metric. And portfolio construction should not assume Bitcoin behaves like an equity because its recent price path resembles one, since its VIX correlation is a fraction of Tesla's and its attribution profile matches neither equity. That low integration is consistent with the diversification evidence reported by Bouri et al. (2017) and with the market-structure survey of Fang

et al. (2022), although the rolling correlations of Figure 2 show the relationship tightening after 2020 and are therefore better read as a time-varying than a fixed property.

7. Limitations and future work

Five limitations bound these conclusions, each pointing to a specific next step. The most consequential is that attribution shares carry no measure of uncertainty: they are point estimates from one fitted model per asset, and the gaps carrying the interpretation, 60.3% against 55.0% and 5.6% against 9.1% and 12.2%, are untested. Bootstrapping the shares over resampled test periods would establish whether the cross-asset ordering is stable, and it is the single highest-value addition to this work. Second, the cross-section is narrow: one cryptoasset and two equities cannot establish whether the attribution gap is a property of the asset classes or of this pairing, and Tesla is a single high-attention name whose sentiment sensitivity may be idiosyncratic. Extending the design to more cryptoassets with different holder bases and to a broader equity cross-section would separate the two. Third, complete-case handling discards information, shrinking sentiment-inclusive Bitcoin blocks from as many as 4,346 observations to as few as 1,670, and the shrinkage is not random with respect to volatility regime, since missing on-chain and sentiment series cluster early in the sample; multiple imputation could recover the lost power and test whether sample composition drives the null. Fourth, hyperparameters were fixed at moderate values across the grid for tractability, so the null is a statement about this configuration rather than about gradient-boosted or recurrent architectures in general; a narrow, heavily tuned study on the best block per asset would show whether it survives a serious model search. Fifth, the equity activity analogue is imperfect, since dollar volume and the Amihud measure are stand-ins for the real activity that on-chain metrics capture directly, and daily frequency may be the wrong resolution, since intraday data could reveal short-horizon predictability around discrete sentiment shocks, and the reflexivity channel is a natural candidate for a study at higher frequency where the mechanical and behavioral components would separate more cleanly.

Funding: none. **Conflicts of interest:** none declared. **Acknowledgements:** none. **Data and code:** available from the author on request. **Reproducibility:** all results derive from a single 315-column merged panel using the walk-forward protocol of Section 3.4, with fixed seeds for the stochastic models and the hyperparameters stated in Section 3.3.

References

- Alabi, Ken. 2017. "Digital Blockchain Networks Appear to Be Following Metcalfe's Law." *Electronic Commerce Research and Applications* 24: 23–29. <https://doi.org/10.1016/j.elerap.2017.06.003>.
- Alternative.me. 2024. "Crypto Fear & Greed Index: Methodology." Accessed 2026. <https://alternative.me/crypto/fear-and-greed-index>.
- Baker, Malcolm, and Jeffrey Wurgler. 2006. "Investor Sentiment and the Cross-Section of Stock Returns." *Journal of Finance* 61 (4): 1645–1680. <https://doi.org/10.1111/j.1540-6261.2006.00885.x>.
- Beckmann, Joscha, Teo Geldner, and Jan Wüstenfeld. 2024. "The Relevance of Media Sentiment for Small and Large Scale Bitcoin Investors." *Journal of International Financial Markets, Institutions and Money* 92: 101963. <https://doi.org/10.1016/j.intfin.2024.101963>.

- Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society, Series B* 57 (1): 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- Bollerslev, Tim. 1986. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31 (3): 307–327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1).
- Bouri, Elie, Peter Molnár, Georges Azzi, David Roubaud, and Lars Ivar Hagfors. 2017. "On the Hedge and Safe Haven Properties of Bitcoin: Is It Really More Than a Diversifier?" *Finance Research Letters* 20: 192–198. <https://doi.org/10.1016/j.frl.2016.09.025>.
- Chen, Tianqi, and Carlos Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. New York: ACM. <https://doi.org/10.1145/2939672.2939785>.
- Cho, Kyunghyun, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. "Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation." In *Proceedings of EMNLP 2014*, 1724–1734. Stroudsburg, PA: ACL. <https://doi.org/10.3115/v1/D14-1179>.
- Da Gama Silva, Paulo Vitor Jordão, Marcelo Cabus Klotzle, Antonio Carlos Figueiredo Pinto, and Leonardo Lima Gomes. 2019. "Herding Behavior and Contagion in the Cryptocurrency Market." *Journal of Behavioral and Experimental Finance* 22: 41–50. <https://doi.org/10.1016/j.jbef.2019.01.006>.
- Diebold, Francis X., and Roberto S. Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics* 13 (3): 253–263. <https://doi.org/10.1080/07350015.1995.10524599>.
- Engle, Robert F. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica* 50 (4): 987–1007. <https://doi.org/10.2307/1912773>.
- Fama, Eugene F. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance* 25 (2): 383–417. <https://doi.org/10.2307/2325486>.
- Fang, Fan, Carmine Ventre, Michail Basios, Leslie Kanthan, David Martinez-Rego, Fan Wu, and Lingbo Li. 2022. "Cryptocurrency Trading: A Comprehensive Survey." *Financial Innovation* 8 (1): 13. <https://doi.org/10.1186/s40854-021-00321-6>.
- Fischer, Thomas, and Christopher Krauss. 2018. "Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions." *European Journal of Operational Research* 270 (2): 654–669. <https://doi.org/10.1016/j.ejor.2017.11.054>.
- Granger, Clive W. J. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica* 37 (3): 424–438. <https://doi.org/10.2307/1912791>.
- Hansen, Peter Reinhard. 2005. "A Test for Superior Predictive Ability." *Journal of Business and Economic Statistics* 23 (4): 365–380. <https://doi.org/10.1198/073500105000000063>.
- Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. "Long Short-Term Memory." *Neural Computation* 9 (8): 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Jegadeesh, Narasimhan, and Sheridan Titman. 1993. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance* 48 (1): 65–91. <https://doi.org/10.1111/j.1540-6261.1993.tb04702.x>.
- Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." *Advances in Neural*

- Information Processing Systems* 30: 3146–3154.
- Khedr, Ahmed M., Ifra Arif, P. V. Pravija Raj, Magdi El-Bannany, Saadat M. Alhashmi, and Meenu Sreedharan. 2021. “Cryptocurrency Price Prediction Using Traditional Statistical and Machine-Learning Techniques: A Survey.” *Intelligent Systems in Accounting, Finance and Management* 28 (1): 3–34. <https://doi.org/10.1002/isaf.1488>.
- Kristoufek, Ladislav. 2013. “BitCoin Meets Google Trends and Wikipedia: Quantifying the Relationship Between Phenomena of the Internet Era.” *Scientific Reports* 3: 3415. <https://doi.org/10.1038/srep03415>.
- Lundberg, Scott M., and Su-In Lee. 2017. “A Unified Approach to Interpreting Model Predictions.” *Advances in Neural Information Processing Systems* 30: 4765–4774.
- Nabipour, Mojtaba, Pooyan Nayyeri, Hamed Jabani, Amir Mosavi, Ely Salwana, and Shahab S. 2020. “Deep Learning for Stock Market Prediction.” *Entropy* 22 (8): 840. <https://doi.org/10.3390/e22080840>.
- Newey, Whitney K., and Kenneth D. West. 1987. “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” *Econometrica* 55 (3): 703–708. <https://doi.org/10.2307/1913610>.
- O’Hara, Maureen. 1995. *Market Microstructure Theory*. Cambridge, MA: Blackwell.
- Shen, Dehua, Andrew Urquhart, and Pengfei Wang. 2019. “Does Twitter Predict Bitcoin?” *Economics Letters* 174: 118–122. <https://doi.org/10.1016/j.econlet.2018.11.007>.
- Tetlock, Paul C. 2007. “Giving Content to Investor Sentiment: The Role of Media in the Stock Market.” *Journal of Finance* 62 (3): 1139–1168. <https://doi.org/10.1111/j.1540-6261.2007.01232.x>.
- Whaley, Robert E. 2000. “The Investor Fear Gauge.” *Journal of Portfolio Management* 26 (3): 12–17. <https://doi.org/10.3905/jpm.2000.319728>.
- White, Halbert. 2000. “A Reality Check for Data Snooping.” *Econometrica* 68 (5): 1097–1126. <https://doi.org/10.1111/1468-0262.00152>.